iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** I **Month of publication:** January 2022

DOI: https://doi.org/10.22214/ijraset.2022.39809

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

Medical Image Fusion Using Deep Learning Mechanism

Nukapeyyi Tanuja
PG Student, Department of Electronics and Communication Engineering, University College of Engineering, JNTUK, Kakinda, Andhra Pradesh, India

Abstract: ***Sparse representation(SR) model named convolutional sparsity based morphological component analysis is introduced for pixel-level medical image fusion. The CS-MCA model can achieve multicomponent and global SRs of source images, by integrating MCA and convolutional sparse representation(CSR) into a unified optimization framework. In the existing method, the CSRs of its gradient and texture components are obtained by the CSMCA model using pre-learned dictionaries. Then for each image component, sparse coefficients of all the source images are merged and then fused component is reconstructed using the corresponding dictionary. In the extension mechanism, we are using deep learning based pyramid decomposition. Now a days deep learning is a very demanding technology. Deep learning is used for image classification, object detection, image segmentation, image restoration.***

Keywords: CNN, CT, MRI, MCA, CS-MCA.

I. INTODUCTION

A. Image Fusion

Image Fusion is a combination of two or more images to get a single précised image,that is more effective with respect to quality analysis and visual quality. The vast majority of current equipment is incapable of reliably transmitting such information. Many data sources can be integrated using image fusion algorithms. The spatial and spectral resolution qualities of the blended image may be complimentary. On the other hand, traditional image fusion approaches can destroy the spectral information of multispectral data when combining.. In satellite imaging, there are two types of images. Satellites transmit panchromatic images at the maximum feasible resolution, whereas multispectral data is conveyed at a lower resolution. This will usually be two to four times lower. At the reception station, the panchromatic signal is received. Image fusion can be done in a variety of ways. The high pass filtering approach is the most basic. The DWT, uniform rational filter bank, and Laplacian pyramid are used in later approaches.

B. Medical Image Fusion

The term "image fusion" has become widely utilised in medical diagnoses and treatment. When a large number of patient photos are gathered and then superimposed or blended to provide extra information, the word is employed. Fused images can be created by combining data from several imaging modalities such as magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), and single photon emission computed tomography (SPECT) (SPECT). These images are used for a variety of applications in radiology and radiation oncology. CT pictures are commonly used to assess tissue density changes, but MRI images are frequently utilised to diagnose brain tumours.

Radiologists must combine data from a variety of imaging types to make a good diagnosis. The use of merged, anatomically consistent images in the diagnosis and treatment of cancer is particularly advantageous. Image fusion software has recently been developed by companies such as Nicesoft, Velocity Medical Solutions, Mirada Medical, Keosys, MIMvista, IKOE, and BrainLAB for both enhanced diagnostic reading and use in conjunction with radiation treatment planning systems. Because diagnostic images can be superimposed on radiation planning images, IMRT target tumour volumes can be calculated more precisely.

C. Magnetic Resonance Imaging (MRI)

MR imaging is a non-invasive imaging technique that creates comprehensive three-dimensional anatomical images. It's frequently used to detect diseases, diagnose them, and track their progress. It is based on cutting-edge technology that excites and detects changes in the rotational axis of protons in the water that makes up biological tissues.

Magnetic resonance imaging (MRI) uses powerful magnets to create a strong magnetic field that compels protons in the body to align with it. The protons are activated and spin out of equilibrium, straining against the magnetic field's pull, when a radiofrequency current is pulsed through the patient. The MRI sensors can detect the energy produced as the protons realign with the magnetic field when the radiofrequency field is switched off. The quantity of energy released and the time it takes for the protons to realign with the magnetic field vary depending on the environment and the chemical makeup of the molecules. On the basis of these magnetic properties, physicians can distinguish between distinct types of tissues.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

D. Computed Tomography (CT)

CT is a non-invasive technology for obtaining CT pictures of every area of the human body without superimposing neighbouring structures. Quantum noise, X-ray scattering by the patient, beam hardening, and nonlinear partial volume effects are only a few of the flaws or picture abnormalities that can occur during X-ray CT measurements. To generate good quality photos for quantitative analysis, image processing with Adobe Photoshop, ImageJ, Corel PHOTO-PAINT, and Origin software was employed. By altering the colours or intensities of a picture, image enhancement techniques can improve the signal-to-noise ratio and highlight image features. CT scans produce two-dimensional images of a "slice" of the body, but the information can also be used to generate three-dimensional views

E. Deep Learning

Deep learning is a machine learning technique for developing artificial intelligence systems. It is based on artificial neural networks (ANNs), which are designed to perform complicated analysis of large amounts of data by passing it through a network of neurons. Deep neural networks are available in a variety of sizes and shapes (DNN). The most common type of neural network used to recognise patterns in photos and video is deep convolutional neural networks (CNN or DCNN). DCNNs originated from ordinary artificial neural networks, which use a threedimensional neural pattern activated by an animal's visual brain. Deep convolutional neural networks are most commonly employed for object recognition, image classification, and recommendation systems, although they are occasionally utilised for herbal language processing.

F. Convolution Neural Networks (CNN)

Deep artificial neural networks known as convolutional neural networks are used to classify images, cluster them according to similarity, and recognise objects within scenes. Faces, street signs, tumours, platy puses, and a range of other visual data are all recognised by them.

The primary building blocks of convolutional neural networks are convolutional layers. Convolution is the basic principle of applying a filter to an input to gain activation. When the same filter is applied to an input multiple times, a feature map is created, illustrating the positions and strength of a recognised feature in an input, such as an image.

The ability of convolutional neural networks to learn a large number of filters in parallel, particular to a training dataset, under the restrictions of a given predictive modelling problem, such as picture categorization, is its distinguishing feature. As a result, input photographs have very precise features that may be recognised from any position.

Convolutional networks use optical character recognition (OCR) to digitise text and enable natural-language processing on analogue and hand-written documents, where the images are symbols to be transcribed. CNNs can be used when sound is represented visually as a spectrogram. Graph convolutional neural networks have recently been utilised to analyse text and graph data using convolutional networks. The efficiency of convolutional nets (ConvNets or CNNs) in picture identification is one of the primary reasons why the world has woken up to the usefulness of deep learning. They are leading major advances in computer vision (CV), which has obvious implications in self-driving cars, robots, drones, security, medical diagnosis, and blindness treatments.

II. THE CNN MODEL FOR MEDICAL IMAGE FUSION

The convolutional network is seen in Figure 5.1. It's a Siamese network with the two branches' weights restricted to the same value. Each branch has three convolutional layers and one max-pooling layer, similar to the network employed in the previous method. We use a considerably lighter model in this work to reduce memory usage and boost computational efficiency by removing a fully linked layer from the network.

After concatenation, the 512 feature maps are directly related to a two-dimensional vector. The minor mode uses just roughly 1.66 MB of physical memory in single precision, according to calculations. Finally, this 2dimensional vector is fed to a 2-way SoftMax layer, which produces a probability distribution over two classes.

The two classes correspond two kinds of normalized weight assignment results, namely, “first patch 1 and second patch 0” and “first patch 0 and second patch 1”, respectively. The probability of each class indicates the possibility of each weight assignment. In this situation, also considering that the sum of two output probabilities is 1, the probability of each class just indicates the weight assigned to its corresponding input patch.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

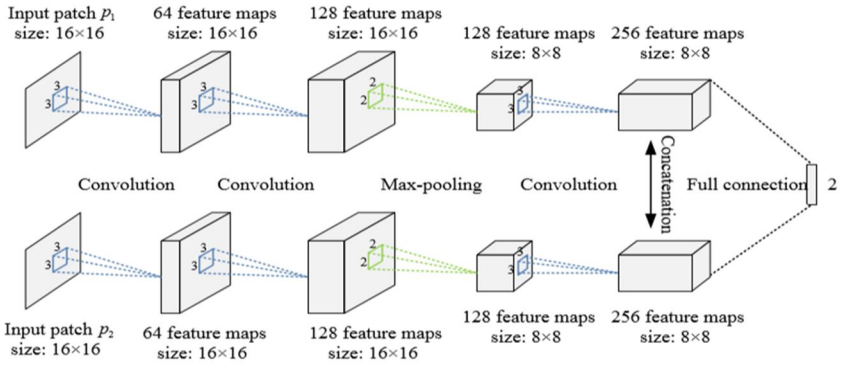


Figure 1. Architecture for CNN training

The network is trained by high-quality image patches and their blurred versions. In the training process, the spatial size of the input patch is set to 16 × 16 according to the analysis. The creation of training examples is based on multi-scale Gaussian filtering and random sampling. The SoftMax loss function is employed as the optimization objective and we adopt the stochastic gradient descent (SGD) algorithm to minimize it. The training process is operated on the popular deep learning framework.

Because the network comprises a fully connected layer with fixed dimensions (pre-defined) on input and output data, the network's input must be fixed in size to ensure that the fully connected layer's input data is fixed. To handle source images of any size in image fusion, break the photos into overlapping patches and feed each patch pair into the network, however this will result in a lot of repetitive calculations. To fix this issue, we first convert the fully connected layer into an equivalent convolutional layer with two 8 8 512 kernels. Following the conversion, the network can analyse source images of any size to create a dense prediction map, with each prediction (a 2-dimensional vector) containing the source image's size

A. Detailed Fusion Scheme

The schematic diagram of the proposed medical image fusion algorithm is shown in Fig. 2. The algorithm can be summarized as the following four steps.

1) *Step 1:* CNN-based weight map generation. Feed the two source images A and B to the two branches of the convolutional network, respectively. The weight map W is generated using the approach described above.

2) *Step 2:* Pyramid decomposition. Decompose each source image into a Laplacian pyramid. Let $\{A\}$ and $\{B\}$ respectively denote the pyramids of A and B, where l indicates the l-th decomposition level. Decompose the weight map W into a Gaussian pyramid $\{W\}$. The total decomposition level of each pyramid is set to the highest possible value$\lfloor \log 2 \min(H,W) \rfloor$, where $H \times W$ is the spatial size of source images and$\lfloor \cdot \rfloor$denotes the flooring operation.

3) *Step 3:* Coefficient fusion. For each decomposition level l, calculate the local energy map (sum of the squares of the coefficients within a small window) of $\{A\}$ and $\{B\}$, respectively.

$$E_A^l(x,y) = \sum_m \sum_n L\{A\}^l(x+m, y+n)^2$$

$$E_B^l(x,y) = \sum_m \sum_n L\{B\}^l(x+m, y+n)^2 \qquad (1)$$

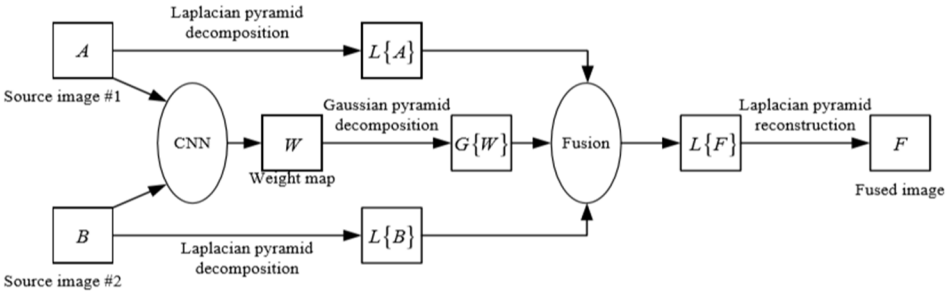


Figure. 2 Proposed medical image fusion

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

The similarity measure used for fusion mode determination is calculated as

$$M^l(x,y) = \frac{2\sum_m \sum_n L\{A\}^l(x+m,y+n)L\{B\}^l(x+m,y+n)}{E_A^l(x,y)+E_B^l(x,y)}. \quad (2)$$

The range of this measure is [-1 1] and a value closer to 1 indicates a higher similarity. A threshold t is set to determine the fusion mode to be used. If $(x,) \geq t$, the "weightedaverage" fusion mode based on the weight map W is adopted as

$$L\{F\}^l(x,y) = G\{W\}^l(x,y) \cdot L\{A\}^l(x,y) + \left(1 - G\{W\}^l(x,y)\right) \cdot L\{B\}^l(x,y). \quad (3)$$

If $(x,) < t$, the "selection" fusion mode via comparing the local energy in Eq. (1) is applied as

$$L\{F\}^l(x,y) = \begin{cases} L\{A\}^l(x,y), \text{ if } E_A^l(x,y) \geq E_B^l(x,y) \\ L\{B\}^l(x,y), \text{ if } E_A^l(x,y) < E_B^l(x,y) \end{cases} \quad (4)$$

The fusion strategy can be summarized as a whole shown in Eq. (5).

Step 4: Laplacian pyramid reconstruction. Reconstruct the fused image F from the Laplacian pyramid $\{F\}$.

$$L\{F\}^l(x,y) = \begin{cases} G\{W\}^l(x,y) \cdot L\{A\}^l(x,y) + \left(1 - G\{W\}^l(x,y)\right) \cdot L\{B\}^l(x,y), \text{ if } M^l(x,y) \geq t \\ L\{A\}^l(x,y), \text{ if } M^l(x,y) < t \ \& \ E_A^l(x,y) \geq E_B^l(x,y) \\ L\{B\}^l(x,y), \text{ if } M^l(x,y) < t \ \& \ E_A^l(x,y) < E_B^l(x,y) \end{cases}$$

III. PYRAMID DECOMPOSITION

Pyramid, also known as pyramid representation, is a multi-scale signal representation developed by the computer vision, image processing, and signal processing fields in which a signal or picture is smoothed and subsampled repeatedly. Scale-space representation and multiresolution analysis are both based on pyramid representation.

A. Pyramid Generation

There are two main types of pyramids: lowpass and bandpass.

A lowpass pyramid is created by smoothing an image using an appropriate smoothing filter and then subsampling it by a factor of two in each coordinate direction. The image that results is subsequently subjected to the same method, and the cycle is repeated several times. Each iteration of this technique yields a smaller image with higher smoothing but lower spatial sampling density (that is, decreased image resolution). The full multi-scale representation will look like a pyramid if depicted visually, with the original image on the bottom and each cycle's smaller image layered one atop the other. To compute pixelwise differences, a bandpass pyramid is created by creating the difference between images at consecutive levels in the pyramid and conducting image interpolation between adjacent levels of resolution.

B. Pyramid Generation Kernels

A variety of smoothing kernels have been proposed for the creation of pyramids. Among the recommendations made, the binomial kernels arising from binomial coefficients stand out as a particularly valuable and philosophically well-founded class. As a result, given a two-dimensional image, we can apply the (normalised) binomial filter (1/4, 1/2, 1/4) twice or more along each spatial dimension and then subsample the picture by a factor of two. To build a compact and efficient multi-scale representation, this procedure can be done as many times as needed.If specified prerequisites are met, intermediate scale levels can be constructed without the subsampling stage, resulting in an oversampled.Because of the increased processing performance of today's CPUs, broader support Gaussian filters can be used as smoothing kernels in the pyramid generation phases in some cases. or hybrid pyramid.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

If specified prerequisites are met, intermediate scale levels can be constructed without the subsampling stage, resulting in an oversampled.Because of the increased processing performance of today's CPUs, broader support Gaussian filters can be used as smoothing kernels in the pyramid generation phases in some cases. or hybrid pyramid.

IV. GAUSSIAN PYRAMID

Using a Gaussian average, the following pictures in a Gaussian pyramid are scaled and weighted down (Gaussian Blur). Each pixel bearing a local average corresponds to a neighbourhood pixel on a lower level of the pyramid. This method is especially handy when creating textures.

The following is how the Gaussian pyramid is calculated. The source image is convolved using a Gaussian kernel. As previously indicated, the final image is a low pass filtered version of the original image. The cut-off frequency can be controlled with this option. The Laplacian is computed using the difference between the original image and the low pass filtered image. This process is done until you get a batch of band-pass filtered photos (since each is the difference between two levels of the Gaussian pyramid). As a result, the Laplacian pyramid is made up of band pass filters.

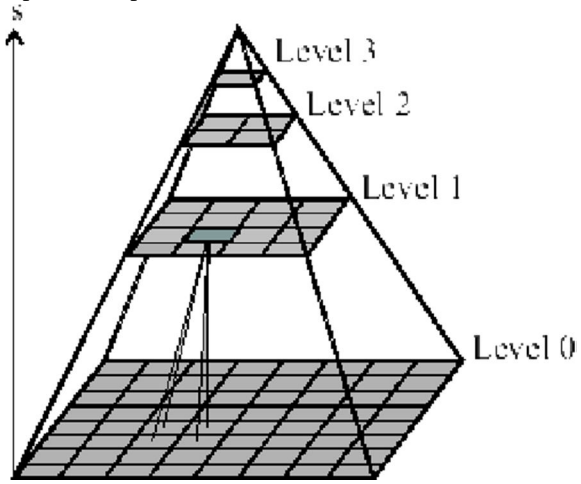


Figure 3: the filtered images stacked one on top of the other form a tapering pyramid structure

Implementation Let $I(x, y)$ be the original image. The Gaussian pyramid on image I is defined as:

$$G_0(x, y) = I$$

$$G_{i+1}(x, y) = REDUCE\big(G_i(x, y)\big)$$

The image is convolved with a Gaussian low pass filter for the REDUCE procedure. The filter mask is set up so that the centre pixel is given greater weight than the surrounding pixels, and the remaining terms are chosen so that their sum equals one. The Gaussian kernel is defined as follows:

$$w(r, c) = w(r)w(c)$$

where, $w(r) = \left(\frac{1}{4} - \frac{a}{2}\ \frac{1}{4}\ \ a\ \ \frac{1}{4}\ \frac{1}{4} - \frac{a}{2}\right)$, a is chosen in the range 0.3 to 0.6. The prediction error $L_0(x, y)$is then given by

$$L_i(x, y) = G_i(x, y) - EXPAND\big(G_{i+1}(x, y)\big)$$

The EXPAND operation is defined as follows:

$$G_{i+1}(x, y) = 4 \sum_{m=-2}^{2} \sum_{n=-2}^{2} w(m, n) G_i\left(\frac{x - m}{2}, \frac{y - n}{2}\right)$$

©IJRASET: All Rights are Reserved

Only terms for which (x-m)/2 and (y-n)/2 are integers are included in the sum. Rather than encode G_0,L_0 and G_1 is encoded. This results in a net data compression because:

1) $L_0$is largely uncorrelated, and so may be represented pixel by pixel with many fewer bitsthanG_0

2) G_1islow pass filtered, and so may be encoded at a reduced sample rate. Further data compression is achieved by iterating this process.

By repeating these steps several times, a sequence of images$L_0, L_1, L_2, \dots, L_n$ are obtained. If we now imagine these images stacked one above another, the result is a tapering pyramid data structure - hence the name. The Laplacian pyramid can thus be used to represent images as a series of band-pass filtered images, each sampled at successively sparser densities.

V. LAPLACIAN PYRAMID

Burt et al. suggested the Laplacian Pyramid (LP) for compact image representation. A Laplacian pyramid is similar to a Gaussian pyramid, but it saves the difference image between each level's blurred versions. To enable reconstruction of the high-resolution image using the difference photos on higher levels, only the smallest level is not a difference image. This method can be used to compress images.

The basic steps of the LP are as follows:

1) Convolve the original image g_0 with a lowpass filter w (e.g., the Gaussian filter) and subsample it by two to create a reduced lowpass version of the image $-g_1$.

2) This image is then up sampled by inserting zeros in between each row and column and interpolating the missing values by convolving it with the same filter w to create the expanded lowpass image g'_1 which is subtracted pixel by pixel from the original to give the detail image $-L_0$ given by

$$L_0=g_0-g'_1$$

In order to achieve compression, rather than encoding g_0, the images L_0 and g_1 are encoded. Since g_1 is the lowpass version of the image it can be encoded at a reduced sampling rate and since L0 is largely decorrelated it can be represented by far fewer bits than required to encode g_0.

The above steps can be performed recursively on the lowpass and subsampled image g_1 a maximum of N number of times if the image size is $2^N \times 2^N$ to achieve further compression. Thus the end result is a number of detail images L_0, L_1, …, L_N and the lowpass image g_N. Each recursively obtained image in the series is smaller in size by a factor of four compared to the previous image and its centre frequency reduced by an octave.

The inverse transform to obtain the original image g_0 from the N detail images L_0, L_1, …, L_N and the lowpass image g_N is as follows:

a) g_N is up sampled by inserting zeros between the sample values and interpolating the missing values by convolving it with the filter w to obtain the image g'_N.

b) The image g'_N is added to the lowest level detail image L_N to obtain the approximation image at the next upper level:

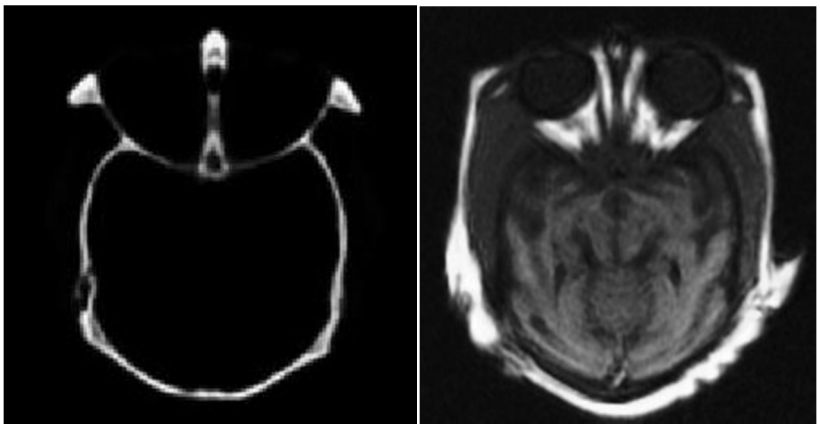
$$g_{N-1} = L_N + g'_N$$

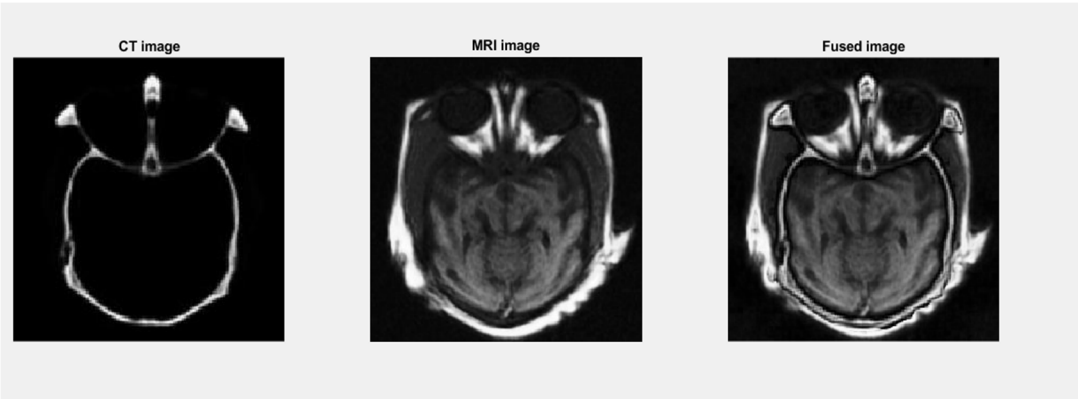
c) Steps 1 and 2 are repeated on the detail images L_0, L_1, …, L_{N-1} to obtain the original image.

VI. RESULTS

All of the experiments were carried out in MATLAB 2016b under high-speed CPU circumstances to reduce running time, as shown in Figure 6.1. Any fusion algorithm's goal is to combine the necessary information from both source photos in the output image. It is impossible to judge a fused image solely by looking at the output image or measuring fusion metrics. It should be evaluated both qualitatively and quantitatively utilising fusion measures. Fusion metrics and image quality assessment (IQA) measures such as peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), correlation coefficient (CC), root mean square error (RMSE), and entropy are used to assess the success of the proposed approach (E). Every fusion algorithm aspires to create a high-resolution fused image. For increased quality, all of these measurements in the fused image should have ideal values. The fusion metric with the greatest value is underlined in bold characters.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |





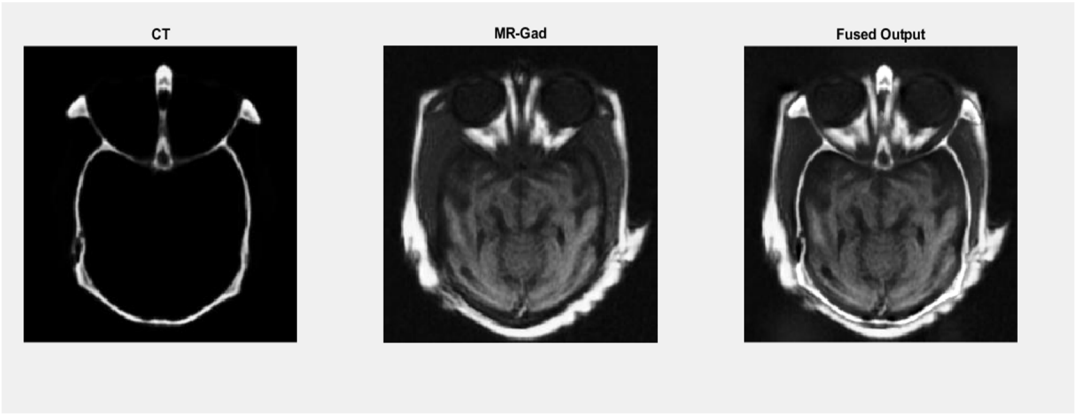


TABLE 6.1: Quantitative Analysis of Fusion Methods for Dataset

Methodology	PSNR (in dB)	RMSE	CC	SSIM	Entropy
existing	58.69	0.31055	0.1236	0.98	6.92
Proposed method	**77.51**	**0.03303**	**0.98974**	**1**	**6.97**

Table 6.1 consists of various fusion metric parameters such as PSNR, RMSE, CC, SSIM and entropy. The best values are highlighted in bold letters. Our proposed method obtained for better values over all the existing fusion methods discussed in the literature

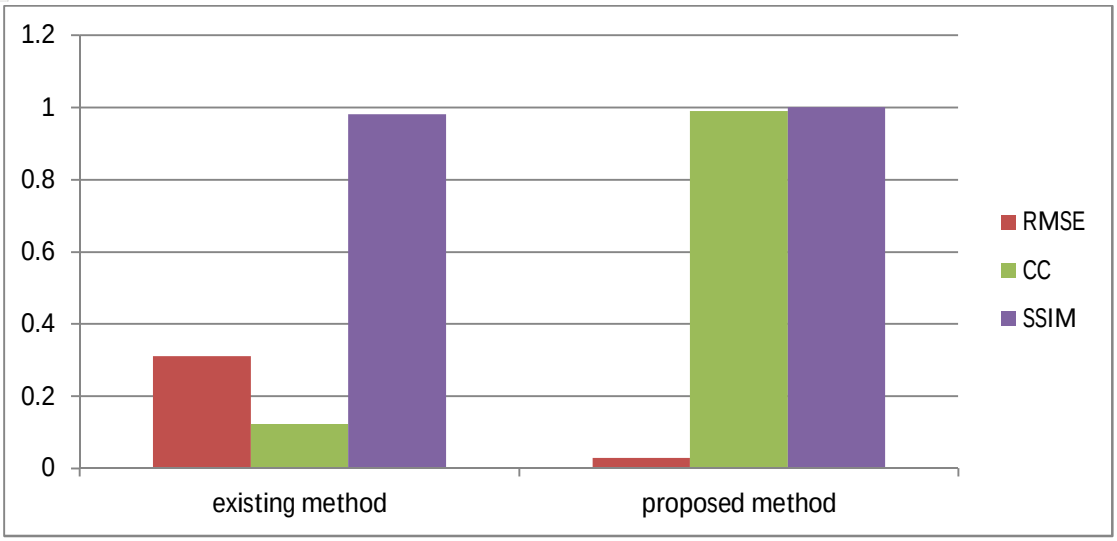


Fig. 6.4. Performance analysis of RMSE,CC,SSIM for dataset with CS-MCA and CNN fusion methodologies.

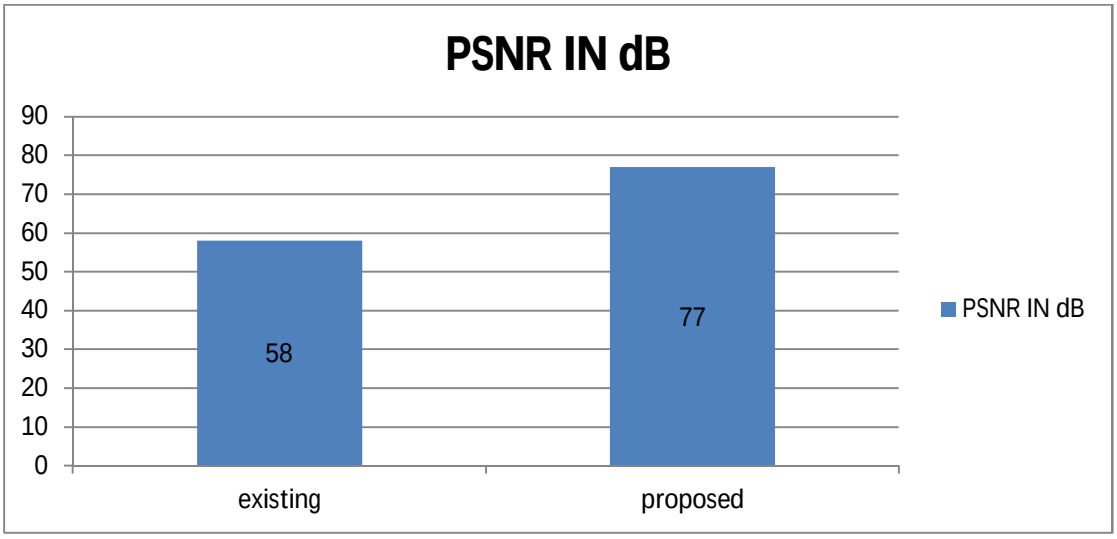


Fig. 6.5Analysis of PSNR performance for dataset using CS-MCA and CNN fusion methods.

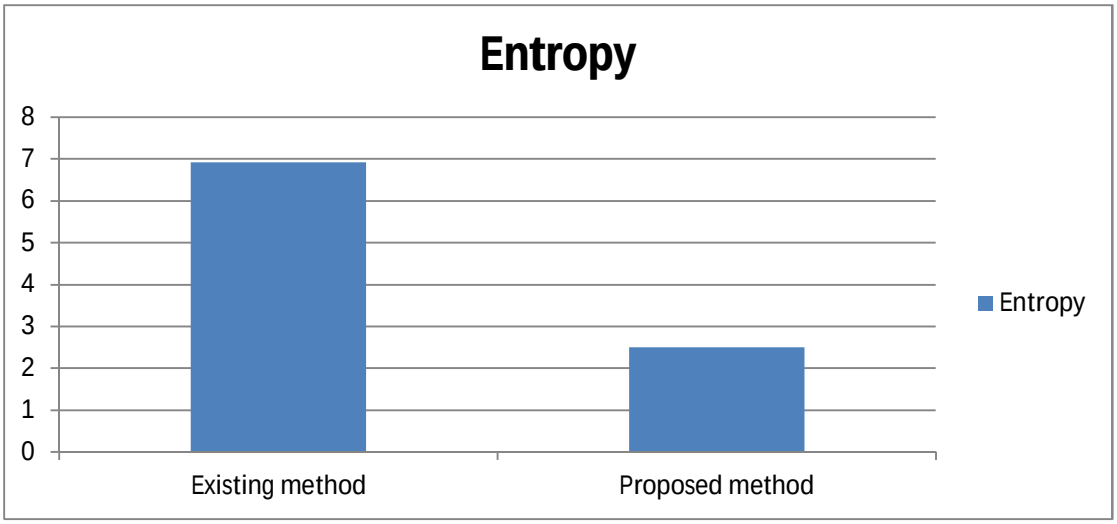


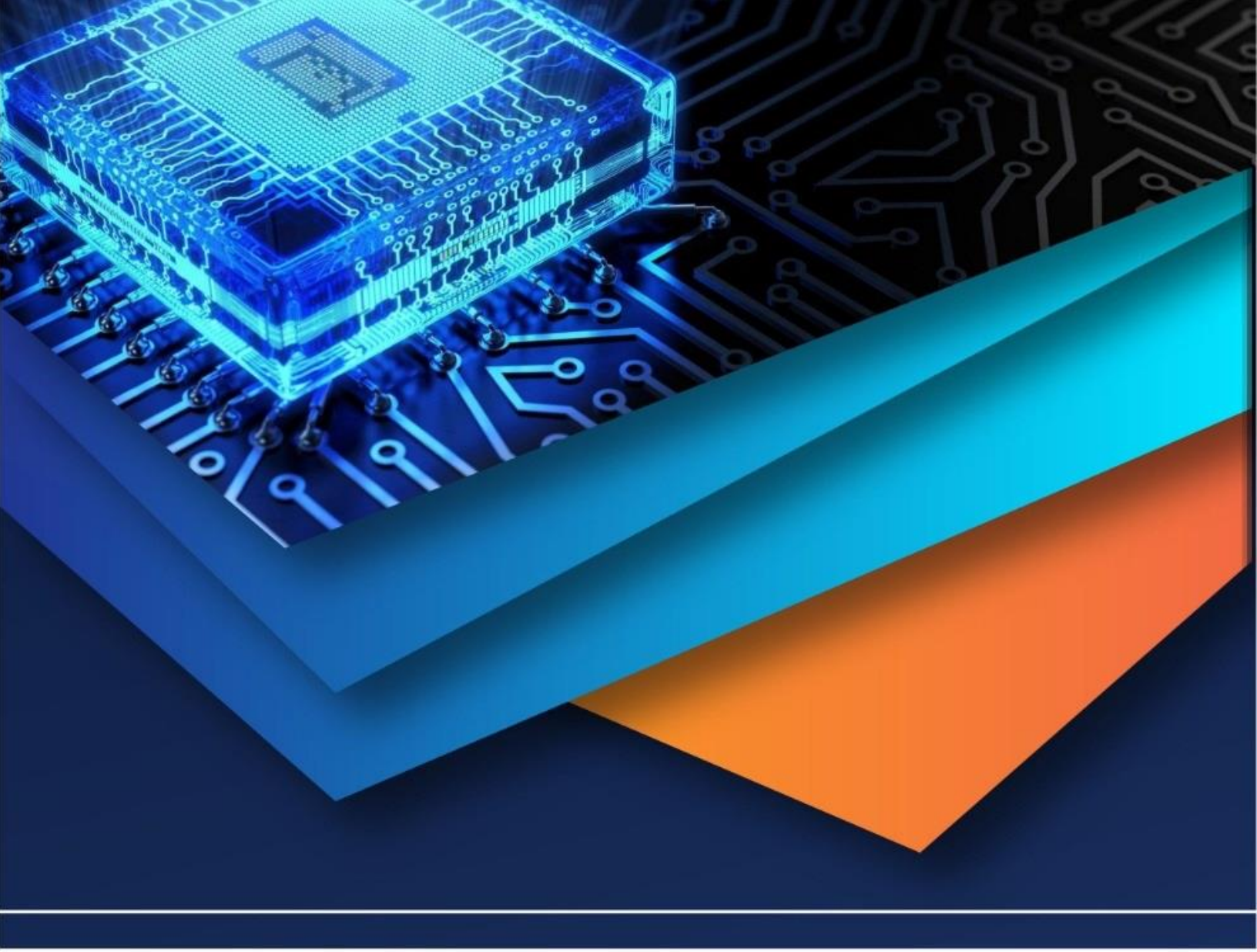
Fig. 6.6 Entropy performance for datasets using CS-MCA and CNN fusion approaches.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

VII. CONCLUSION

This paper proposes a medical image fusion method based on convolutional neural networks. To construct a direct mapping from source images to a weight map including the integrated pixel activity information, we use a Siamese network. The key uniqueness of this approach is that it can use network learning to combine activity level measurement and weight assignment, overcoming the barrier of artificial design. Some prominent image fusion techniques, such as multi-scale processing and adaptive fusion mode selection, are used to generate perceptually good results. The proposed method can produce high-quality outcomes in terms of visual quality and objective metrics, according to the findings of the experiments.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)